N· D'ORDRE : 12

THÈSES

PRÉSENTÉES

A LA FACULTÉ DES SCIENCES DE L'UNIVERSITÉ DE BESANÇON

POUR OBTENIR

LE GRADE DE DOCTEUR ÈS SCIENCES MATHÉMATIQUES

PAR

Jean-Philippe MASSONIE
Licencié-ès-Sciences

PREMIÈRE THÈSE

Utilisation d'une statistique linéaire, correcte pour estimer la vitesse d'évasion des étoiles dans le plan galactique

DEUXIÈME THÈSE

L'axiomatique du calcul des probabilités, par Kolmologoroff

Soutenues le 30 octobre 1964 devant la Commission d'examen

MM.	F. NAHON	Président
	J. DELAHAYE	Examinateurs
	J. L. RIGAL	
	W. FRICKE	
	Ch. FEHRENBACH	
	D. DUGUÉ	

N· D'ORDRE : 12

THÈSES

PRÉSENTÉES

A LA FACULTÉ DES SCIENCES DE L'UNIVERSITÉ DE BESANÇON

POUR OBTENIR

LE GRADE DE DOCTEUR ÈS SCIENCES MATHÉMATIQUES

PAR

Jean-Philippe MASSONIE
Licencié-ès-Sciences

PREMIÈRE THÈSE

Utilisation d'une statistique linéaire, correcte pour estimer la vitesse d'évasion des étoiles dans le plan galactique

DEUXIÈME THÈSE

L'axiomatique du calcul des probabilités, par Kolmologoroff

Soutenues le 30 octobre 1964 devant la Commission d'examen

MM.	F. NAHON	Président
	J. DELAHAYE	Examinateurs
	J. L. RIGAL	
	W. FRICKE	
	Ch. FEHRENBACH	
	D. DUGUÉ	

FACULTÉ DES SCIENCES DE L'UNIVERSITÉ DE BESANÇON

Doyen Honoraire
M. GLANGEAUD

Professeurs Honoraires
MM. BAILLAUD, DUFFIEUX, TRILLAT, UEBERSFELD, VINCENSINI.

Maître de Conférences Honoraire
M. GRANIER

Doyen

M.	JACQUEMAIN	Chimie Appliquée

Assesseur

M.	MESNAGE	Chronométrie

UNIVERSITEITSBIBLIOTHEEK - LEIDEN -

Professeurs

MM.	TRONCHET	Botanique
	CHATELET	Mathématiques
	PERROT	Chimie Générale
	DREYFUSS	Géologie Minéralogie
	THEOBALD	Géologie
	CROISOT	Méthodes Mathématiques de la Physique
	DELHAYE	Astronomie
	QUANTIN	Botanique
	CHALEAT	Mécanique Appliquée
	RIPPLINGER	Physiologie Animale
	NAHON	Mécanique Théorique
	BERNARD	Chimie Physique
	RIGAL	Mathématiques
	VIENOT	Physique Générale
Mme.	ROZIS	Physique
M.	THIEBAUT	Géologie Minéralogie

Professeurs sans chaire

MM.	POTIER	Cristallographie Physique
	DEVIN	Chimie Générale
	GOMOT	Zoologie
	MOTCHANE	Electronique

Maîtres de conférences

MM.	VALLEE	Mécanique Appliquée
	REAL	Zoologie
	GOUARNE	Mathématiques
Mlle	LEGRAND	Botanique
M.	THEOBALD	Physique PCB. SPCN
	REMOND	Chargé d'Enseignement de la Maîtrise de Conférences de Chronométrie

Conseiller administratif universitaire
M. LAMBERT

A mes parents.

A Monsieur le Professeur NAHON.

J'ai commencé ce travail à l'observatoire de Marseille où Monsieur le Professeur Fehrenbach m'avait accueilli et donné les possibilités matérielles pour l'entreprendre. Je tiens ici à l'en remercier vivement.

Je suis heureux d'exprimer ma reconnaissance à Monsieur le Professeur Delhaye qui m'a permis par la suite de m'installer à l'observatoire de Besançon.

Je remercie Monsieur le Professeur Dugué d'avoir bien voulu s'intéresser à ce travail, Monsieur le Professeur Rigal qui m'a ouvert les portes de son laboratoire de Calcul Numérique où a été réalisée toute la partie "application" de ma thèse.

Je suis heureux d'exprimer ma gratitude à Monsieur le Professeur Fricke qui m'a fait bénéficier au Rechen - Institut d'Heidelberg de son expérience des problèmes dynamiques et cinématiques.

Je remercie aussi tous ceux qui m'ont aidé dans le travail matériel de préparation de cette thèse.

Enfin, je tiens à remercier tout spécialement Monsieur le Professeur Nahon qui m'a enseigné ce que je sais de dynamique et de statistique stellaire, qui m'a dirigé tout au long de ce travail et qui a su être pour moi un Maître, tant du point de vue scientifique que du point de vue humain.

Publi. Inst. Statist. Univ. Paris
14 (1965), 1-64

UTILISATION D'UNE STATISTIQUE LINÉAIRE CORRECTE POUR ESTIMER LA VITESSE D'ÉVASION DES ÉTOILES DANS LE PLAN GALACTIQUE

MASSONIE J.-P.

INTRODUCTION

En 1928, dans le B. A. N. (1), Oort étudiant la distribution des vitesses d'étoiles, constatait dans la direction de la rotation galactique l'absence de vitesses supérieures à 63 km/s. Construisant une courbe de densité des étoiles de vitesses supérieures à 63 km/s, en fonction de la longitude, il trouvait un trou dans la direction 1 = 55° (anciennes coordonnées).

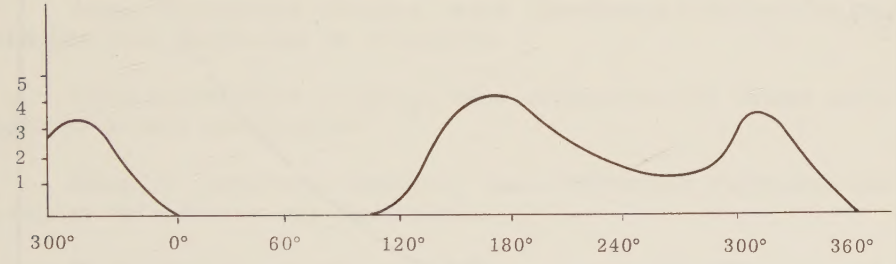

Figure 1 - Distribution des points de grande vitesse en longitude galactique d'après Oort.

Si on rapporte les vitesses au centre absolu des vitesses, on s'aperçoit qu'elles sont toutes contenues dans un grand cercle de rayon R. Si V_c est la vitesse circulaire au voisinage du soleil, on a

$$R = V_c + 63 \qquad \text{(fig. 2)}$$

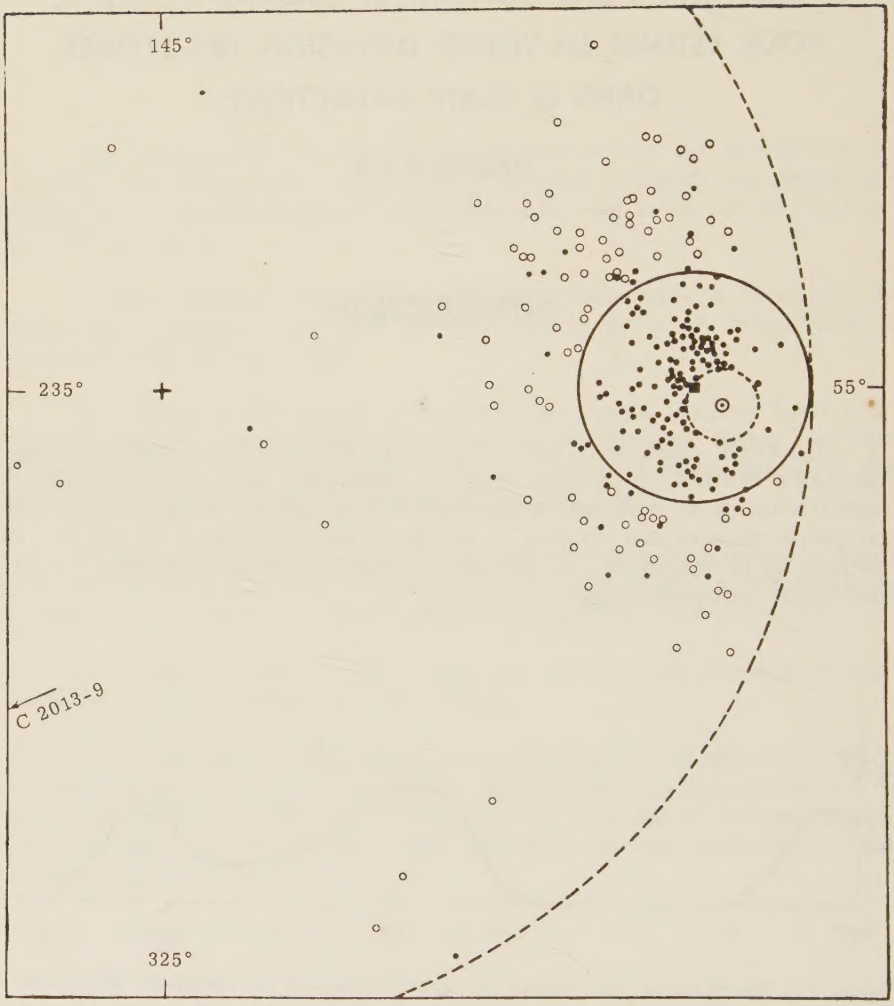

Figure 2 - Distribution des grandes vitesses dans le plan galactique.

Nous nous proposons de reprendre ce problème d'un point de vue statistique en tenant compte du fait que s'il existe une troncature, elle doit être un peu au delà des plus grandes vitesses. Notre but est donc de déterminer plus exactement les caractéristiques du grand cercle signalé par Oort.

Nous savons que lorsque l'énergie cinétique d'une étoile dépasse le potentiel du point où elle se trouve, cette étoile s'échappe. Le grand cercle correspond donc à la limite à ne pas dépasser par les vitesses, sinon il y aurait évasion.

Fricke (2) signale qu'il faut se servir des plus grandes vitesses pour estimer la vitesse d'évasion et suggère que cette vitesse peut être nettement supérieure aux plus grandes vitesses observées.

Nous voyons donc apparaître deux problèmes, l'un dynamique : quel modèle choisir, quelle forme donner au potentiel, comment relier le potentiel et la distribution des vitesses ; l'autre statistique : comment estimer une troncature connaissant les plus grandes valeurs d'une distribution.

Il restera pour en arriver aux applications à résoudre le problème du matériel de vitesses à utiliser.

Nous verrons donc dans un premier chapitre l'étude du problème dynamique.

Dans un deuxième chapitre, nous aborderons l'estimation statistique d'un paramètre de troncature.

Dans le troisième chapitre, nous estimerons les autres paramètres de nos distributions.

Dans le quatrième chapitre, nous étudierons l'influence des erreurs de mesures sur l'estimateur.

Enfin dans le cinquième chapitre, nous appliquerons nos méthodes à différents matériels de vitesses radiales.

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41547625_0034

CHAPITRE I

PROBLÈME DYNAMIQUE ET CINÉMATIQUE

Il nous faut un modèle de distribution des vitesses ; un tel modèle ne peut être obtenu que par des considérations cinématiques sur les vitesses d'étoiles connues, et dynamiques sur la structure même de la galaxie.

L'hypothèse de départ sera que la galaxie est un système de masse finie et que ce système est stationnaire. Moyennant cette hypothèse, on sait qu'il ne peut pas y avoir d'étoiles dont la vitesse soit supérieure en un lieu donné à une certaine valeur que nous appellerons vitesse d'évasion V_e.

Cette vitesse V_e est reliée au potentiel du lieu par

$$V_e = \sqrt{2\Omega}$$

où Ω est le potentiel.

Pour les étoiles de faible vitesse (inférieure à 60 km/s), la répartition ellipsoïdale, encore qu'elle présente quelques difficultés, est à peu près satisfaisante.

Cependant, dès lors qu'on aborde les grandes vitesses, elle ne rend plus compte des observations.

Fricke (2) propose un nouveau modèle à partir de considérations dynamiques que nous allons rappeler.

Soit à trouver

$$f(x, y, z, u, v, w, t)\, dx\, dy\, dz\, du\, dv\, dw = \mathbf{P}$$

qui est la probabilité qu'une étoile ait ses coordonnées et sa vitesse tel que

$$x < X < x + dx \qquad u < U < u + du$$

$$y < Y < y + dy \qquad v < V < v + dv$$

$$z < Z < z + dz \qquad w < W < w + dw$$

Les hypothèses de Fricke sur le système sont les suivantes

a) stationnaire

b) masse finie

c) symétrie axiale

d) conservation du moment angulaire

e) égalité des masses des étoiles.

Montrant les difficultés d'une fonction de la forme Gram-Charlier pour les calculs ultérieurs, il utilise un développement de forme polynomiale généralisée.

$$f(E_1, E_2) = \sum_{i,k} \alpha_{ik} E_1^i E_2^k$$

où α_{ik} sont des constantes, i et k non forcément des entiers,

$E_1 = 2\Omega(R, z) - V^2$ intégrale de l'énergie

$E_2 = Rv$ moment angulaire

$V^2 = u^2 + v^2 + w^2$

u, v, w les composantes des vitesses dans le système d'axe suivant v est orienté dans la direction de la rotation, u est orienté en sens inverse du centre galactique, w est l'axe perpendiculaire au plan galactique de façon à former avec v et u un trièdre direct.

Nous savons que la densité est liée à la fonction de densité des vitesses par

$$\rho(R, z) = \iiint_{0 \leq v \leq \sqrt{2\Omega}} f(E_1, E_2)\, du\, dv\, dw$$

On a donc

$$\rho(R, z) = \sum_{i,k} \alpha_{ik} \iiint_{0 \leq v \leq \sqrt{2\Omega}} (2\Omega - V^2)^i \, (Rv)^k \, du \, dv \, dw$$

Nous passerons sur le détail, on obtient

$$\rho(R, z) = 2\pi \sum_{i,k} \frac{\alpha_{i,2k}}{2k+1} \beta \left(k + \frac{3}{2}, \; i + 1\right) R^{2k} \, (2\Omega)^{i+k+\frac{3}{2}}$$

avec $i > 0$

On voit disparaître les termes impairs des puissances de E_2.

Pour que les hypothèses de départ soient respectées, nous devons imposer à cette fonction de vérifier l'équation de Poisson. D'autre part, la masse étant finie, le potentiel doit être équivalent à $\frac{M}{r}$ quand r tend vers l'infini.

En deuxième approximation, on prendra

$$\Omega = \frac{M}{r} + \frac{M_2(\mu)}{r^{-2}} \qquad \text{où : } \mu = \cos \chi$$

avec

$$R = r \sin \chi$$
$$r^2 = R^2 + z^2$$
$$z = r \cos \chi$$

En portant cette valeur du potentiel dans l'équation de Poisson et en identifiant avec le ρ obtenu plus haut, on montre que l'on doit avoir la condition

$$i \geqslant \frac{5}{2} + k$$

où i indique l'exposant de plus bas de E_1.

La disparition des termes d'exposant impair de k présente une difficulté ; en effet à une fonction de densité donnée correspond pour la densité des vitesses une fonction

$$f = f_1 + f_2$$

où

$$f_1 = E_1^{\frac{5}{2}} \sum_{k=0 \text{ à } n, l} \alpha_{k+l, 2k} \, E_1^{k+l} \cdot E_2^{2k}$$

et $$f_2 = \sum_{\substack{k \text{ de } 0 \text{ à } n \\ l \text{ de } 0 \text{ à } n}} \alpha_{i,2k+1} E_1^i E_2^{2k+1}$$

Fricke montre que pour f_1 le terme de plus bas degré pour E_1 est en $\frac{5}{2}$, la partie f_2 ne peut pas être atteinte et l'on est obligé de s'en tenir à des suppositions quant à son terme de plus bas degré en E_1.

La fonction f_2 est impaire, donc

$$f_2(-v) = -f_2(v)$$

pour que $f = f_1 + f_2 \geqslant 0$, il faut $f_1 \geqslant |f_2|$ pour tous les points de l'espace des phases.

Comme $f_1 \simeq E_1^{k+l+\frac{5}{2}} E_2^{2k}$ pour $r \longrightarrow \infty$

autrement dit

$$f_1 \simeq r^{-l-\frac{5}{2}} \longrightarrow 0$$

f_2 doit avoir un terme de plus bas degré au moins égal à celui de f_1.

Par d'autres considérations pour la limite $r = 0$, Fricke montre que le premier exposant de f_2 est de la forme

$$f_2 = E_1^{\frac{5}{2}+N+\frac{l}{2}}$$

Nous devons voir maintenant ce que donne cette fonction pour les vitesses radiales que nous utiliserons seulement comme étant meilleures.

Soit donc $f(u, v, w,)$ la distribution des vitesses.

Nous savons que la vitesse radiale est la projection de la vitesse sur la direction de l'étoile.

Soit R la vitesse radiale

T la composante tangentielle dans le plan galactique

W est toujours la composante perpendiculaire au plan galactique.

Alors

$$u = u(R, T ; l)$$

$$v = v(R, T ; l)$$

et $f(u, v, w)$ devient $k(R, T, W ; l) = f((u(R, T ; l) ; v(R, T ; l) ; w))$

La fonction de distribution des vitesses radiales sera donc obtenue en intégrant par rapport à T et w

$$h(R) = \iint_{\pi} k(R, T, w ; l)\, d\,T\, dw$$

où π est le plan perpendiculaire en son extrémité au vecteur R.

Nous voyons qu'un point de vitesse radiale peut être la projection de n'importe quel point du plan π.

Donc le grand cercle déterminant la vitesse d'évasion se transforme en une cardioïde (cf. figure ci-jointe).

Cependant si théoriquement le problème est simple, nous allons voir que pratiquement il reste des difficultés.

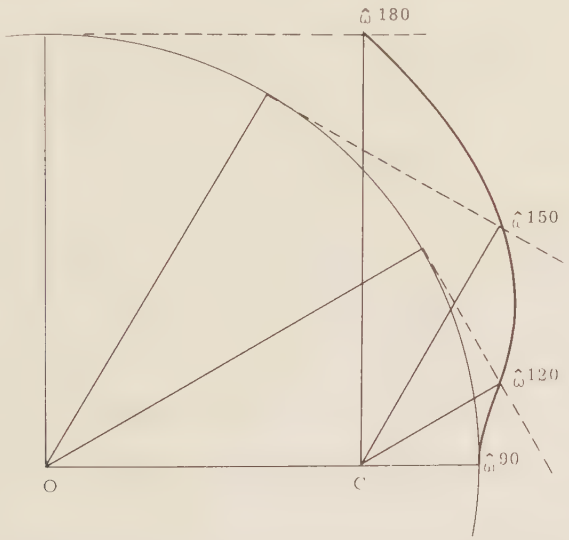

Figure 3 - Construction de la cardioïde, limite d'évasion pour les vitesses radiales.
O : point de vitesse absolue nulle ; C : point représentant la vitesse circulaire au voisinage du soleil ; $\hat{\omega}(l)$: vitesse radiale maximum des étoiles de longitude l.

En effet, si E_1 reste identique dans sa forme quand on passe de (u, v, w) à (R, T, w), E_2 introduit des sin l et cos l qui vont compliquer le développement.

D'autre part, nous verrons par la suite que nous aurons à prendre la fonction inverse de la fonction de distribution, il nous faudra donc une expression relativement simple.

Puisque nous nous intéressons aux vitesses proches de la vitesse d'évasion, nous constatons que le terme de plus bas degré est prépondérant, donc nous nous limiterons à ce terme ; mais il faudra faire très attention, quand on passera aux applications, à se limiter à des vitesses proches de la troncature.

Il reste une dernière difficulté : c'est l'origine des vitesses. Le modèle de Fricke est centré au zéro des vitesses, or nous avons des vitesses radiales par rapport au soleil ou au centroïde local, et nous ne voulons pas introduire la vitesse de rotation du soleil qui est mal connue.

Nous allons donc étudier le terme de plus bas degré du modèle de Fricke pour voir ce qu'il donne au point de vue de sa forme uniquement.

Soit donc

$$f(u, v, w) \simeq \alpha(\omega^2 - V^2)^{\frac{5}{2}} \qquad \text{où } \omega^2 = 2\Omega$$

$$f = \alpha(\omega^2 - (u^2 + v^2 + w^2))^{\frac{5}{2}}$$

$$= \alpha(\omega^2 - (R^2 + T^2 + w^2))^{\frac{5}{2}}$$

En intégrant par rapport à w et T, il vient

$$h(R) = A(\omega^2 - R^2)^{\frac{7}{2}} \qquad \text{où } A \text{ est une constante.}$$

Si R est proche de ω on peut écrire

$$h(R) \simeq A(2\omega)^{\frac{7}{2}} (\omega - R)^{\frac{7}{2}}$$

qui donne finalement pour la fonction de répartition

$$H(R) \simeq 1 - B(\omega - R)^{\frac{9}{2}}$$

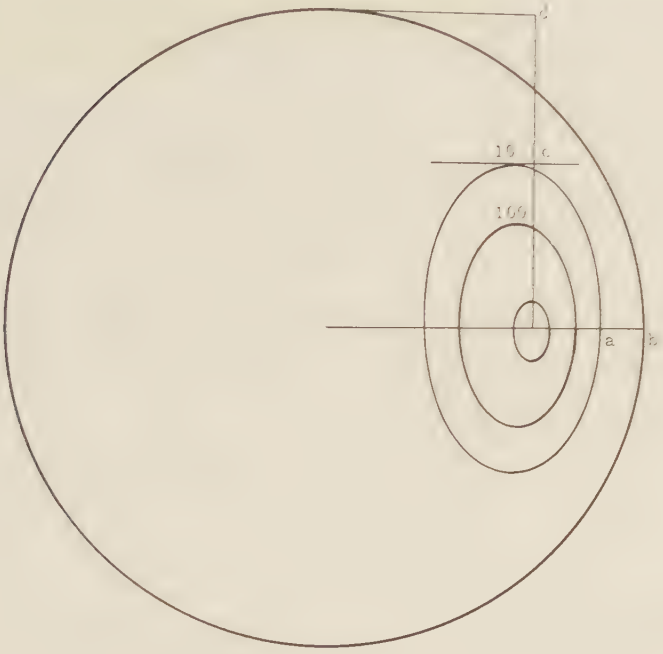

Figure 4 - On voit sur cette figure que la zone comprise entre la région de forte densité et ω, est très différente suivant la longitude.
Les deux cas extrêmes : pour l = 90 (a, b)
pour l = 180 (c, d)

que nous mettons sous la forme :

$$H(R) = 1 - \left(\frac{\omega - R}{b}\right)^{\frac{9}{2}}$$

ou de manière plus générale

$$H(R) = 1 - \left(\frac{\omega - R}{b}\right)^{k}$$

Nous verrons que nous trouverons des estimateurs simples pour les paramètres ω et b ; par contre k sera difficile à estimer. Nous pourrons donc prendre $k = 9/2$.

Le modèle que nous avons adopté peut être contesté, nous voulons donc donner maintenant deux justifications quant à sa forme.

Si nous reprenons le problème de la relation entre la fonction de densité des vitesses f et la densité des étoiles ρ, nous savons les difficultés rencontrées par Camm en particulier. Il est intéressant de développer les fonctions f et ρ, en série car on trouvera entre les termes correspondants des développements, des relations intégrales du même type que celle existant entre f et ρ.

Cherchons donc un développement de f suivant les puissances de E_2, tel que les coefficients de ce développement seront des fonctions de densité pour E_1.

Autrement dit

$$f(E_1, E_2) = \sum \alpha_k E_2^k f_k(E_1)$$

avec

$$f_k(E_1) > 0$$

$$\int f_k(E_1) \, d E_1 = 1$$

On s'aperçoit que si en outre on cherche un système de fonctions f_k orthonormées, il n'y a en général pas de solution. Par contre, si l'on se contente de chercher un système complet, des solutions sont possibles. Il est alors normal de penser à un système de puissance.

En prenant pour f_k des puissances de E_1, on s'aperçoit que d'une part on a un développement en puissance de E_1 qui fournit donc un système complet, au sens de la décomposition des fonctions ; et d'autre part, en choisissant convenablement les coefficients α, on aura bien des fonctions de densité pour E_1. En effet, $E_1 > 0$ donc aussi les puissances de E_1 et on peut choisir les α de façon que

$$\int_0^{\sqrt{2\Omega}} E_1^k \, d E_1 = 1$$

Le deuxième point de vue qui justifie un tel développement est d'ordre statistique.

Dans le cas d'une fonction tronquée sans faire d'autres hypothèses sur la forme de la fonction, pour travailler sur les valeurs extrêmes et trouver leurs distributions asymptotiques, Von Mises (3) (11) utilise un développement à proximité de la troncature.

On a $$F(\omega) = 1$$

et $$f(\omega) = 0$$

On suppose que

$$f'(\omega) = \ldots = f^{(k-2)}(\omega) = 0$$

et que

$$f^{(k-1)}(\omega) = F^{(k)}(\omega) = (-1)^{k-1}\, b$$

k étant un entier supérieur à 1

On suppose en outre que

$|f^{(k)}(x)| \leqslant A$ pour x tel $\omega - \delta \leqslant x \leqslant \omega$ δ petit.

Le développement de Taylor donne

$$F(x) = F(\omega) + \frac{(x - \omega)^k}{k\,!}\, F^{(k)}(\omega) + \ldots$$

Autrement dit

$$F(x) \simeq 1 + \frac{(x - \omega)^k}{k\,!}\, b\, (-1)^{k-1}$$

ou $$F(x) \simeq 1 - B\,(\omega - x)^k$$

avec $$B = \frac{b}{k\,!}$$

On voit donc apparaître la forme que nous avons adoptée.

Remarque

Nous avons montré que, à la distribution de Fricke, correspond pour la fonction de distribution F des vitesses radiales, un exposant $k = 9/2$ pour le terme de plus bas degré.

Nous avons tabulé différentes fonctions, les unes en gardant la forme

$$F = 1 - \left(\frac{v^2 - R^2}{d^2}\right)^{k'}$$

les autres

$$F = 1 - \left(\frac{v - R}{c}\right)^k$$

On s'aperçoit alors que l'approximation avec $1 - \left(\frac{\omega - R}{c}\right)^k$ de la fonction $1 - \left(\frac{\omega^2 - R^2}{d^2}\right)^{k'}$, est bien meilleure et sur un intervalle plus important si l'on prend

$$k = k' - 1$$

Nous ferons donc les estimations en prenant $k = \frac{9}{2} - 1 = \frac{7}{2}$ et nous pensons que c'est la meilleure valeur qui corresponde au modèle de Fricke.

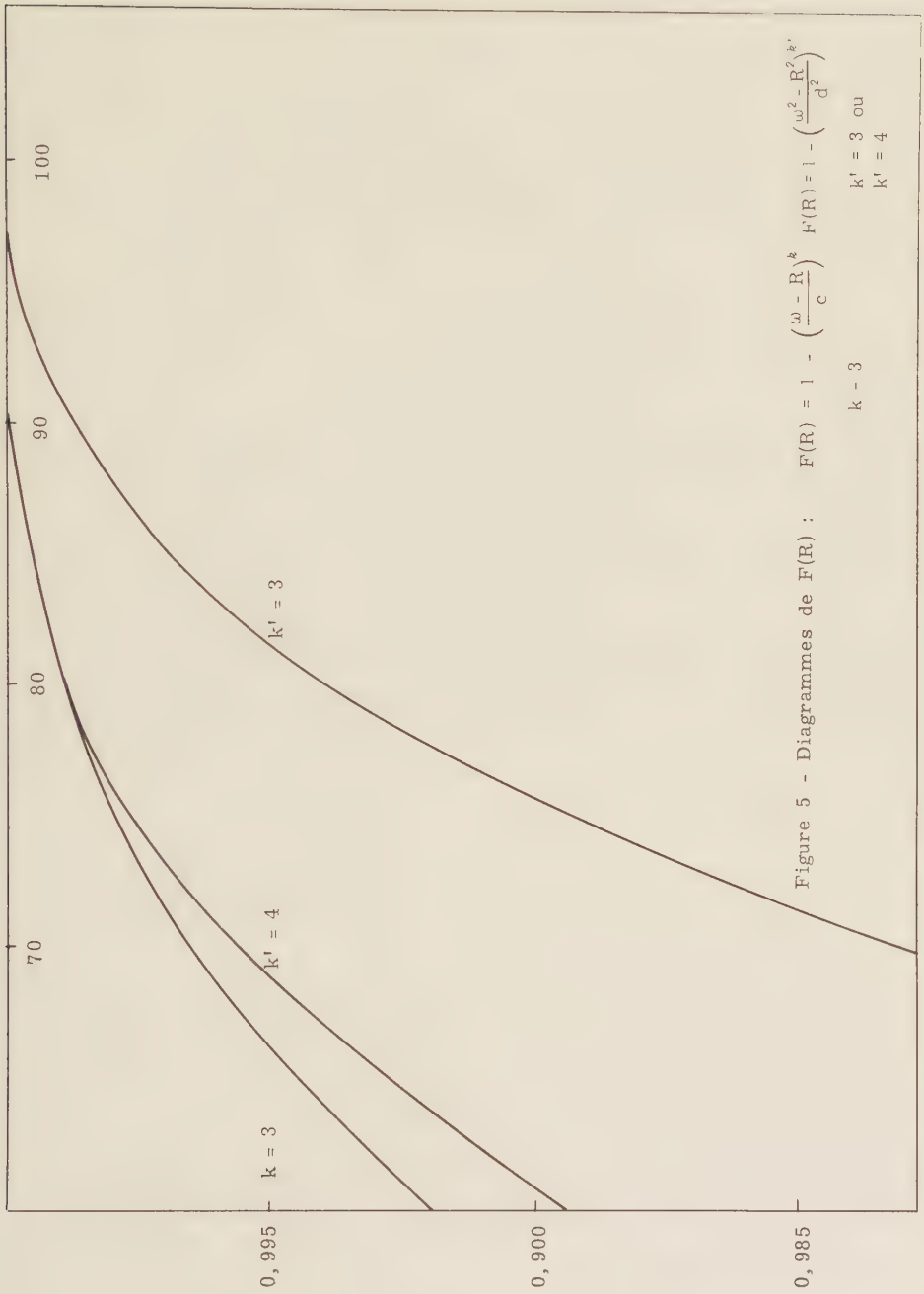

Figure 5 - Diagrammes de F(R) : $F(R) = 1 - \left(\frac{\omega - R}{c}\right)^k$ $F'(R) = 1 - \left(\frac{\omega^2 - R^2}{d^2}\right)^{k'}$

k = 3 k' = 3 ou k' = 4

CHAPITRE II

ESTIMATION D'UN PARAMÈTRE DE TRONCATURE

Nous avons vu que le problème qui se pose à nous est l'estimation du paramètre de troncature d'une fonction de distribution.

Nous allons tout d'abord rappeler un résultat simple sur ce problème dans le cas d'une distribution uniforme.

a) *Loi uniforme.*

Soit un échantillon de taille n, distribué uniformément entre 0 et ω. La densité de probabilité est donc $\frac{1}{\omega}$ et la vraisemblance d'un tel échantillon s'écrit

$$L \equiv f(x_1 ; \omega) \, . \, f(x_2 ; \omega) \, \ldots \, f(x_n ; \omega) = \frac{dx_1 \, \ldots \, dx_n}{\omega^n}$$

Soit M_n la plus grande valeur, on a

$$P_r(M_n < x) = \left(\frac{x}{\omega}\right)^n$$

et la densité correspondante est

$$\frac{n(x)^{n-1}}{\omega^n} \, dx$$

Si on applique le maximum de vraisemblance à L, donnée plus haut, on s'aperçoit que la plus grande valeur M_n donne une estimation de ω. Mais si nous calculons la moyenne de la distribution de M_n, nous trouvons

$$E\,M_n = \int_o^{\omega} M_n \cdot n \frac{M_n^{n-1}}{\omega^n} d\,M_n = \frac{n\omega}{n+1}$$

$$= \omega - \frac{\omega}{n+1}$$

Nous avons donc une estimation qui n'est pas correcte, par contre on peut montrer que la variance tend vers 0 avec n tendant vers l'infini.

Donc estimation : asymptotiquement correcte et efficace.

De même, on peut montrer que l'estimateur M_n est exhaustif. En effet

$L(x_1 \ldots . x_n)$ peut s'écrire

$$L = n \frac{M_n^{n-1}}{\omega^n} \cdot \frac{\partial M_n}{\partial x_1} \cdot \frac{dx_1 \ldots . dx_n}{n\,M_n^{n-1} \frac{\partial M_n}{\partial x_1}}$$

qui est la forme canonique pour une estimation exhaustive (4).

Partant de ce schéma rapide, nous allons appliquer le maximum de vraisemblance, puis nous chercherons successivement à rendre les estimateurs utilisés corrects, efficaces.

b) *Maximum de vraisemblance.*

Nous allons estimer le paramètre de troncature d'une fonction du type de Fricke, c'est-à-dire

$$F(x) = 1 - \left(\frac{\omega - x}{b}\right)^k \qquad \text{pour} \quad x < \omega$$

$$F(x) = 1 \qquad \text{pour} \quad x \geqslant \omega$$

Pour cela nous utiliserons deux valeurs extrêmes, en effet si l'on essaye avec une valeur extrême, de mettre la vraisemblance sous forme canonique, on s'aperçoit que l'estimation n'est plus exhaustive. Nous verrons par la suite si nous pouvons utiliser plus de deux valeurs.

Calculons la probabilité pour que

$$x < X_m < x + dx \qquad \text{et} \qquad y < Y_p < y + dy$$

où X_m est la m° valeur en partant du sommet, et Y_p la p°, avec $m > p$.

On a alors le schéma suivant

$$\underset{n-m \quad X_m \quad m-p-1 \quad Y_p \quad p-1 \quad \omega}{\overset{\alpha \qquad \beta \qquad \gamma}{\longleftarrow\!\!\!

$$\xi \left[\frac{m - 1}{\xi} - \left(1 - \frac{m}{n}\right) \frac{1}{1 - \frac{\xi}{n}} \right] = 1$$

qui donne pour la racine $\hat{\xi}$

$$\hat{\xi} = m$$

puis $$\left(\gamma + 1 - \frac{\beta\theta}{1 - \theta}\right)^k = \frac{1 + \theta^{1/k}}{1 - \theta^{1/k}}$$

dont la racine unique est telle que

$$M = \frac{\hat{\omega} - y}{\hat{\omega} - x}$$

qui nous donne donc l'estimateur

$$\hat{\omega} = \frac{y - Mx}{1 - M}$$

Le maximum du vraisemblance a donc donné un résultat simple dans sa forme finale. Mais l'application n'en sera pas très facile ; pour faire l'étude de l'estimateur, nous rencontrons une difficulté, le paramètre à estimer est une des bornes de l'intégrale

$$\int f(x, \omega)dx = 1$$

donc : $\sqrt{n}(\hat{\omega} - \omega)$ ne suivra pas une loi de Laplace-Gauss.

Cependant nous allons garder la forme simple du résultat : l'estimateur sera une combinaison linéaire.

c) *Estimateur linéaire correct.*

Nous allons donc prendre comme estimateur

$$\hat{\omega} = \frac{y - Mx}{1 - M}$$

et nous lui imposerons d'être correct, autrement dit

$$E\hat{\omega} = \omega$$

Calculons donc $\hat{\omega}$ en fonction des variables ξ et ϑ introduites plus haut.

$$\frac{\xi}{n} = \left(\frac{\omega - x}{b}\right)^k$$

d'où $$x = \omega - \left(\frac{\xi}{n}\right)^h . b \qquad \text{avec} \quad h = \frac{1}{k}$$

$$\frac{\vartheta\xi}{n} = \left(\frac{\omega - y}{b}\right)^k$$

d'où $$y = \omega - \left(\frac{\xi\vartheta}{n}\right)^h b$$

donc $$\hat{\omega} = \frac{1}{1 - M}\left(\omega - \left(\frac{\xi\vartheta}{n}\right)^h b\right) - \frac{M}{1 - M}\left(\omega - \left(\frac{\xi}{n}\right)^h b\right)$$

$$\hat{\omega} = \omega - \left(\frac{\xi}{n}\right)^h b . \frac{\vartheta^h - M}{1 - M}$$

$$E\,\hat{\omega} = \omega - b\,E\left(\frac{\xi}{n}\right)^h . \frac{E(\vartheta^h) - M}{1 - M}$$

La condition $E\,\hat{\omega} = \omega$ impose donc

$$E\,\vartheta^h = M$$

car $E\left(\frac{\xi}{n}\right)^h$ est différent de 0.

La condition de correction nous permet donc de déterminer l'estimateur sans ambiguïté.

Nous pouvons nous demander si le fait de n'utiliser que deux valeurs, n'entraîne pas une perte importante d'informations.

Nous allons donc essayer d'utiliser plusieurs valeurs extrêmes. Le même calcul que plus haut fournira une solution.

On peut aussi utiliser une méthode soulignée par Blom (6) qui considère les valeurs d'un échantillon comme transformées de variables distribuées uniformément entre 0 et 1 ; la transformation étant la fonction inverse de la fonction de distribution initiale.

Blom fournit même des formules d'approximation pour calculer simplement les moments d'ordre 1 et 2 de ces variables.

Une telle méthode nous conduira à un estimateur

$$\hat{\omega} = \sum a_i x_i$$

mais la condition de correction ne nous fournit que deux relations pour déterminer les a_i, ce qui est insuffisant.

Nous pouvons donc imposer à l'estimateur des conditions plus restrictives, qui en même temps permettront de le déterminer sans ambiguïté.

d) *Estimateur linéaire correct et consistant.*

Nous posons toujours le problème de l'estimation du paramètre de troncature de la fonction

$$\begin{cases} F(x) = 1 - \left(\frac{\omega - x}{b}\right)^k & \text{pour} \quad |x| \leqslant \omega \\ = 1 & \text{pour} \quad |x| > \omega \end{cases}$$

et nous utilisons un estimateur linéaire

$$\hat{\omega} = \sum_{i=1}^{p} a_i x_i$$

où x_i est la i° variable en partant du sommet.

Nous imposerons deux conditions

$$\begin{cases} 1) \; E\,\hat{\omega} = \omega \\ 2) \; \text{var}\,(\hat{\omega}) \text{ minimum} \end{cases}$$

La condition (1) impose comme nous avons vu

$$E\,\hat{\omega} = \omega = \sum_{i=1}^{p} a_i E\,x_i$$

La condition (2) impose

$$\text{var}(\hat{\omega}) = \sum_i \sum_j C_{ij}\, a_i\, a_j \quad \text{minimum}$$

avec

$$C_{ij} = \text{Cov}\,(x_i, x_j)$$

et

$$C_{ii} = \text{var}\,(x_i)$$

Pour faciliter les notations, nous poserons

$$E\, x_i = E_i$$

On a donc à rendre minimum la forme quadratique

$$V = \sum_i \sum_j a_i\, a_j\, C_{ij}$$

en tenant compte de la liaison

$$\sum a_i\, E_i = \omega$$

Nous savons que pour avoir la solution de ce problème (7), nous devons annuler les dérivées partielles de la fonction auxiliaire

$$G = \sum_i \sum_j a_i\, a_j\, C_{ij} + \lambda \sum_i \left(a_i\, E_i - \frac{\omega}{n} \right)$$

En fait, nous verrons par la suite que E_i sera de la forme

$$E_i = \omega - D\, F_i$$

où D est une constante et F_i une fonction du rang i.

On est donc obliger d'introduire deux inconnues auxiliaires et l'on aura comme fonction

$$G = \sum_i \sum_j a_i\, a_j\, C_{ij} + \lambda \sum \left(a_i - \frac{1}{n} \right) - \mu \sum a_i\, F_i\, D$$

En dérivant par rapport à a_i, on obtient

$$\sum_j a_j\, c_{ij} + \lambda\, \omega - \mu\, D\, F_i = 0$$

On aura autant de relations de cette forme que d'inconnues a_i.

On a en plus l'identité

$$\sum a_i\, E_i \equiv \omega$$

ou

$$\sum a_i\, (\omega - D\, F_i) \equiv \omega$$

qui donne deux relations

$$\begin{cases} \sum a_i = 1 \\ \sum a_i F_i = 0 \end{cases}$$

Donc si nous prenons une combinaison linéaire d'ordre p comme estimateur, on a

$p + 2$ relations et p inconnues a_i, plus les deux inconnues auxiliaires λ et μ.

Pour que le problème soit résolu, il faut maintenant calculer les variances, covariances et moyennes. Nous avons deux procédés à notre disposition

- calcul direct

- approximations de Blom.

Blom fait l'étude des transformations qui permettent de passer de la distribution uniforme à n'importe quelle autre distribution. En effet, soit X une variable aléatoire de fonction de répartition $F(x)$ et soit $x_1, x_2 \ldots x_n$ un échantillon de taille n.

Posons

$$u_i = F(x_i)$$

Nous faisons ainsi correspondre à chaque valeur x_i, une valeur u_i comprise entre 0 et 1 ; la transformation inverse F^{-1} fait passer de u_i à x_i.

Nous poserons, pour faciliter les choses

$$G \equiv F^{-1}$$

donc

$$x_i = G(u_i)$$

Si nous ordonnons notre échantillon x_i, de même en vertu de la monotonie de F, l'échantillon u_i sera ordonné. Nous pouvons calculer la moyenne, la variance d'une variable u_i. Nous donnons les résultats sans démonstration

fonction densité : $\beta_i(u_i) = \dfrac{n\,!}{(i-1)\,!\;(n-i)\,!}\; u_i^{i-1}\,(1-u_i)^{n-i}$

$$\text{moyenne} \quad : \; p_i = \frac{i}{n+1}$$

$$\text{variance} \quad : \; v_i = \frac{p_i (1 - p_i)}{n+2}$$

$$\text{covariance} \; : \; c_{ij} = \frac{p_i (1 - p_j)}{n+2} \qquad i \leqslant j$$

Blom montre que sous certaines hypothèses, on peut revenir par la transformation G aux moyennes, variances et covariances des variables x_i. On a

$$E\, x_i = G\left(\frac{i}{n+1}\right) + R_i \qquad \text{où } |R_i| < \frac{M}{n}$$

M étant une constante ne dépendant pas de i et n

$$\operatorname{cov}(x_i, x_j) = \frac{1}{n+2}\, p_i(1 - p_j)\, G'(p_i)\, G'(p_j) + R_{ij}$$

avec $|R_{ij}| < \frac{M}{n^2}$ où G' est la dérivée de G.

Avec certaines classes de transformations, on est obligé d'utiliser des formules plus compliquées.

Nous pourrions donc utiliser cette méthode.

Dans le cas qui nous intéresse, nous avons préféré un calcul direct, mais nous avons tenu à signaler cette méthode de Blom qui permet dans certains cas d'obtenir des résultats plus facilement.

Calcul direct des $E(x_i)$

Dans ce paragraphe et les suivants, l'indice i est le rang compté à partir du sommet.

$$E_i \equiv E\, x_i = \int_{-\omega}^{+\omega} x_i \, d\, H(x_i)$$

où H est la fonction de répartition de la i° valeur extrême à partir du sommet.

Passons à la variable ξ_i du changement de variable de Cramer

$$\xi_i = n[1 - F(x_i)]$$

$$E_i = \int_o^h x_i(\xi) \cdot h(\xi) \cdot d\xi$$

or avec la forme adoptée pour F

$$x_i(\xi) = \omega - b \left(\frac{\xi}{n}\right)^h$$

donc $$E_i = \int_o^n \left(\omega - b \left(\frac{\xi}{n}\right)^h\right) \cdot h(\xi) \cdot d\xi$$

et asymptotiquement, on a

$$E_i = \int_o^n \left(\omega - b \left(\frac{\xi}{n}\right)^h\right) \frac{e^{-\xi}\ \xi^{i-1}}{\Gamma(i)} \cdot d\xi$$

qui donne

$$E_i = \omega - \frac{b}{n^h} \frac{\Gamma(i+h)}{\Gamma(i)}$$

en négligeant les infiniment petits d'ordre supérieur à $\frac{1}{n^h}$

Calcul des C_{ij}

De la même manière, on peut calculer les variances

$$C_{ii} = E(x_i - E_i)^2 = \int_o^n (x_i(\xi) - E_i)^2 \cdot \frac{e^{-\xi}\ \xi^{i-1}}{\Gamma(i)} \cdot d\xi$$

$$= \int_o^n \left[\frac{b}{n^h} \frac{\Gamma(i+h)}{\Gamma(i)} - b \left(\frac{\xi}{n}\right)^h\right]^2 \frac{e^{-\xi}\ \xi^{i-1}}{\Gamma(i)} \cdot d\xi$$

ou finalement

$$C_{ii} = \frac{b^2}{n^{2h}} \left[\frac{\Gamma(i+2h)}{\Gamma(i)} - \left(\frac{\Gamma(i+h)}{\Gamma(i)}\right)^2\right]$$

Calcul des C_{ij} ($i \neq j$ covariance)

$$\mathrm{Cov}(x_i\ x_j) \equiv C_{ij} = \iint_{\text{pour } x_i < x_j} x_i(\xi) \cdot x_j(\xi\ \theta) \cdot h(\xi, \theta)\ d\xi \cdot d\theta$$

En utilisant la fonction $h(\xi, \theta)$ définie plus haut, on trouve alors par un calcul identique au précédent

$$C_{ij} = \frac{b^2}{n^{2h}} \left[\frac{\Gamma(i+2h)}{\Gamma(j)} \cdot \frac{\Gamma(j+h)}{\Gamma(i+h)} - \frac{\Gamma(i+h)\,\Gamma(j+h)}{\Gamma(j)\,\Gamma(i)} \right]$$

Le calcul est valable pour $j < i$, par symétrie on a les autres covariances.

Nous sommes maintenant en possession de tous les coefficients.

Remarquons que notre estimateur est indépendant de b, indépendant de n puisque nous avons utilisé des fonctions asymptotiques pour les ξ_i.

Par contre, l'exposant k du terme de plus bas degré du développement de Fricke intervient considérablement.

Remarque.

Le système d'équations linéaires que nous obtenons a une forme très particulière. On a

$$\vec{AM} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}$$

et la matrice M se décompose comme suit

$$M = \begin{pmatrix} \boxed{T} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} \\ C_{ij} & \boxed{T^*} \end{pmatrix}$$

où T^* est la transposée de T et C_{ij} une matrice symétrique. (cf. tableau).

Enfin, en raison de la forme particulière du second membre, les a_i apparaissent comme les termes de la première colonne de la matrice inverse de la matrice M.

Donnons pour terminer, l'expression de la variance de notre estimateur

$$V = \sum_i \sum_j a_i \, a_j \, C_{ij}$$

Remarquons que cette variance dépendra de b et sera de la forme

$$V = \frac{B}{n^{2h}}$$

où $$B = b^2 \sum_i \sum_j a_i \, a_j \, D_{ij}$$

où $$D_{ij} = n^{2h} \, C_{ij}$$

Matrice des variances et covariances pour k = 7/2,
à un facteur $\left(\frac{b^2}{n^{2h}}\right) \cdot 10^{-4}$ près.

1	2	3	4	5	6	7	8	9	10
814	481	356	288	245	215	193	176	162	151
481	619	457	370	315	277	249	227	209	194
356	457	523	423	360	317	284	259	239	222
288	370	423	463	395	347	311	284	261	243
245	315	360	395	423	372	333	304	280	261
215	277	317	347	372	393	352	321	296	276
193	249	284	311	333	352	369	336	310	289
176	227	259	284	304	321	336	350	323	289
162	209	239	261	280	296	310	323	334	311
151	194	222	243	261	276	289	300	311	321

Matrice des variances et covariances pour k = 9/2,
à un facteur $\left(\frac{b^2}{n^2}\right)$. 10^{-4} près.

1	2	3	4	5	6	7	8	9	10
529	288	204	160	133	115	101	90	82	75
288	352	249	196	163	140	123	111	100	92
204	249	277	218	181	156	137	123	112	102
160	196	218	234	194	167	147	132	120	110
133	163	181	194	205	176	155	139	126	116
115	140	156	167	176	184	162	145	132	121
101	123	137	147	155	162	168	151	137	126
90	111	123	132	139	145	151	156	141	130
82	100	112	120	126	132	137	141	145	133
75	92	102	110	116	121	126	130	133	137

Tableau des coefficients de l'estimateur linéaire de ω.

Combinaison d'ordre 10

Paramètre K	2,5	3,0	3,5	4,0	4,5
a_1	1,0680	1,0783	1,1070	1,1402	1,1751
a_2	0,2496	0,3281	0,3912	0,4548	0,5321
a_3	0,1771	0,2377	0,3008	0,3708	0,3964
a_4	0,1464	0,2041	0,2425	0,3022	0,3575
a_5	0,1147	0,1621	0,2383	0,2616	0,3575
a_6	0,1043	0,1733	0,2014	0,3641	0,2083
a_7	0,0910	0,1230	0,2281	0,0008	0,3638
a_8	0,0868	0,0970	0,0945	0,3746	0,3416
a_9	0,0635	0,1478	0,1903	0,2137	0,0511
a_{10}	-1,1015	-1,5515	-1,9942	-2,4829	-2,8164

Combinaison d'ordre 8

Paramètre K	2,5	3,0	3,5	4,0	4,5
a_1	1,190922	1,225821	1,269244	1,328307	1,373236
a_2	0,277836	0,369518	0,458969	0,528034	0,632180
a_3	0,199386	0,271835	0,334667	0,411945	0,487041
a_4	0,159139	0,220050	0,303833	0,409591	0,451740
a_5	0,124816	0,199661	0,268821	0,300605	0,406646
a_6	0,122497	0,190269	0,224250	0,345245	0,228286
a_7	0,111785	0,130661	0,204940	0,064575	0,411773
a_8	-1,186386	-1,607818	-2,064727	-2,388305	-2,990905

Nous avons fait le calcul de la forme quadratique correspondante qui servira pour le calcul de la variance.

$$Q = \sum_i \sum_j a_i \, a_j \, C_{ij}$$

Pour $k = 4,5$

$$Q = 0,396$$

Pour $k = 3,5$

$$Q = 0,472$$

CHAPITRE III

ESTIMATION DES PARAMÈTRES k ET b

A - **Estimation du paramètre k**

Nous avions montré dans une publication précédente que lorsque k augmentait, la valeur donnée par l'estimateur augmentait.

Nous donnons ici deux essais pour estimer k, les deux se soldant par des difficultés d'ordre pratique.

Etant donné la forme de

$$E\, x_i = \omega - b\, E \left(\frac{\xi}{n}\right)^k$$

pour avoir un estimateur indépendant de ω et b, on trouve deux solutions

- une fonction homographique conduisant à un estimateur

$$\frac{x_i - x_j}{x_k - x_1}$$

Dans ce cas, la taille de l'échantillon n'intervient pas.

- une fonction homographique faisant intervenir des valeurs de deux échantillons différents

$$\frac{x_i - x_j}{y_i - y_j}$$

1er estimateur

Supposons deux échantillons ordonnés, indépendants, de taille n et p.

Prenons les variables de rang i et j dans chacun des échantillons et formons

$$\frac{x_i - x_j}{y_i - y_j}$$

Cette quantité tend asymptotiquement vers la quantité $\left(\frac{p}{n}\right)^h$.

En effet

$$\frac{x_i - x_j}{y_i - y_j} = \frac{\xi_j^h - \xi_i^h \left(\frac{b}{n}\right)^h}{\eta_j^h - \eta_i^h \left(\frac{b}{p}\right)^h} \longrightarrow \left(\frac{p}{n}\right)^h$$

Donc si p et n sont assez différents, on doit avoir des résultats sensibles à h. L'application au matériel que nous avions, n'a pas donné de résultat.

2e estimateur

Nous partons de x_1, x_2, x_i et formons la statistique

$$\frac{x_1 - x_2}{x_1 - x_i}$$

Cherchons $h(x_1, x_2, x_i)$, fonction densité jointe de x_1, x_2, x_i.

$$h(x_1, x_2, x_i) = \frac{n!}{(n-i)!\,(i-3)!} (F(x_i))^{i-3} (F(x_2) - F(x_i))^{i-3} \; dF(x_1) \cdot dF(x_2) \cdot dF(x_3)$$

Faisons le changement de variable habituel de Cramer.
Il vient

$$g(\xi_1, \xi_2, \xi_i) = \frac{n!}{(n-i)!\,(i-3)!} \frac{1}{n^i} \left(1 - \frac{\xi_i}{n}\right)^{n-i} (\xi_i - \xi_2)^{i-3} \, d\xi_1 \, d\xi_2 \, d\xi_i$$

avec $$\xi_i = n[1 - F(x_i)]$$

Posons $$\theta_1 = \frac{\xi_1}{\xi_i} \qquad \theta_2 = \frac{\xi_2}{\xi_i}$$

Il vient

$$k(\xi_i, \theta_1, \theta_2) = Q \left(1 - \frac{\xi_i}{n}\right)^{n-i} \cdot \xi_i^{i-3} (1 - \theta_2)^{i-3} \, d\xi_i \, d\theta_1 \, d\theta_2 \cdot J$$

où J est le jacobien de la transformation.

On montre que

$$I - \xi_i^2$$

Donc quand $n \longrightarrow \infty$

$$k(\xi_i, \theta_1, \theta_2) \longrightarrow \frac{1}{\Gamma(i-2)} e^{-\xi} \xi_i^{i-1} (1-\theta_2)^{i-3} \, d\xi_i \, d\theta_1 \, d\theta_2$$

Alors

$$E\left(\frac{x_1 - x_2}{x_1 - x_i}\right) = \iiint \frac{x_1 - x_2}{x_1 - x_i} k(\xi_i, \theta_1, \theta_2) \, d\xi_i \, d\theta_1 \, d\theta_2$$

où x_1, x_2, x_i sont exprimés en fonction de $\xi_i, \theta_1, \theta_2$.

L'intégration par rapport à ξ_i est sans difficulté et il reste

$$E\left(\frac{x_1 - x_2}{x_1 - x_i}\right) = \int_0^1 (1-\theta_2) \, d\theta_2 \int_0^{\theta_2} \frac{\theta_2^h - \theta_1^h}{1 - \theta_1^h} \, d\theta_1$$

En effet

$$x_1 > x_2 > x_i$$
$$0 < \xi_1 < \xi_2 < \xi_i < n$$
$$0 < \theta_1 < \theta_2 < 1$$

Le calcul donne

$$E = h((i-3\,!) \sum_{j=1}^{\infty} \frac{1}{[(j-1)(h+1)]\ldots(j\,h+1)\ldots(j\,h+i-1)}$$

Cette expression a deux inconvénients

1/ pour $i = 2$ elle n'a pas de sens

2/ elle est difficile à manier

Nous avons alors utilisé l'idée suivante

pour $n \longrightarrow \infty$

$F(x_1)$ se "stabilise" à $1 - \frac{1}{n}$

$F(x_2)$ se "stabilise" à $1 - \frac{2}{n}$

$F(x_i)$ se "stabilise" à $1 - \frac{i}{n}$

donc

$$\xi_i \simeq \left(\frac{i}{n}\right) n = i$$

$$\xi_1 \simeq \left(\frac{1}{n}\right) n = 1$$

$$\xi_2 \simeq \left(\frac{2}{n}\right) n = 2$$

Notre statistique devient donc très simple et

$$E\left(\frac{x_1 - x_2}{x_1 - x_i}\right) \simeq \frac{2^h - 1}{i^h - 1}$$

Nous donnons ci-joint les résultats pour $\neq$ valeurs de h, jusqu'à i = 10.

i \ k	2	3	4	5	6	1
2	1	1	1	1		1
3	0,5658	0,5877	0,5986	0,6051	0,6094	0,5
4	0,4142	0,4425	0,4568	0,4654	0,4711	0,33
5	0,3351	0,3661	0,3820	0,3916	0,3980	0,25
6	0,2858	0,3181	0,3348	0,3450	0,3519	0,20
7	0,2517	0,2847	0,3020	0,3125	0,3197	0,18
8	0,2265	0,2599	0,2775	0,2883	0,2956	0,14
9	0,2071	0,2406	0,2585	0,2695	0,2769	0,125
10	0,1916	0,2251	0,2431	0,2542	0,2618	0,11

Nous constatons immédiatement que les valeurs sont très tassées et si l'on compare à ce que donne des échantillons, on s'aperçoit que notre statistique ne donne pas de renseignement, la fluctuation des résultats pratiques étant trop grande. cf. fig. 8, p. 35 .

Par contre, il est encourageant de constater qu'à partir de i = 5 ou 6, pour les différents échantillons utilisés, on avait une excellente convergence vers les valeurs théoriques. Ceci implique donc que le modèle simplifié adopté représente assez correctement les observations.

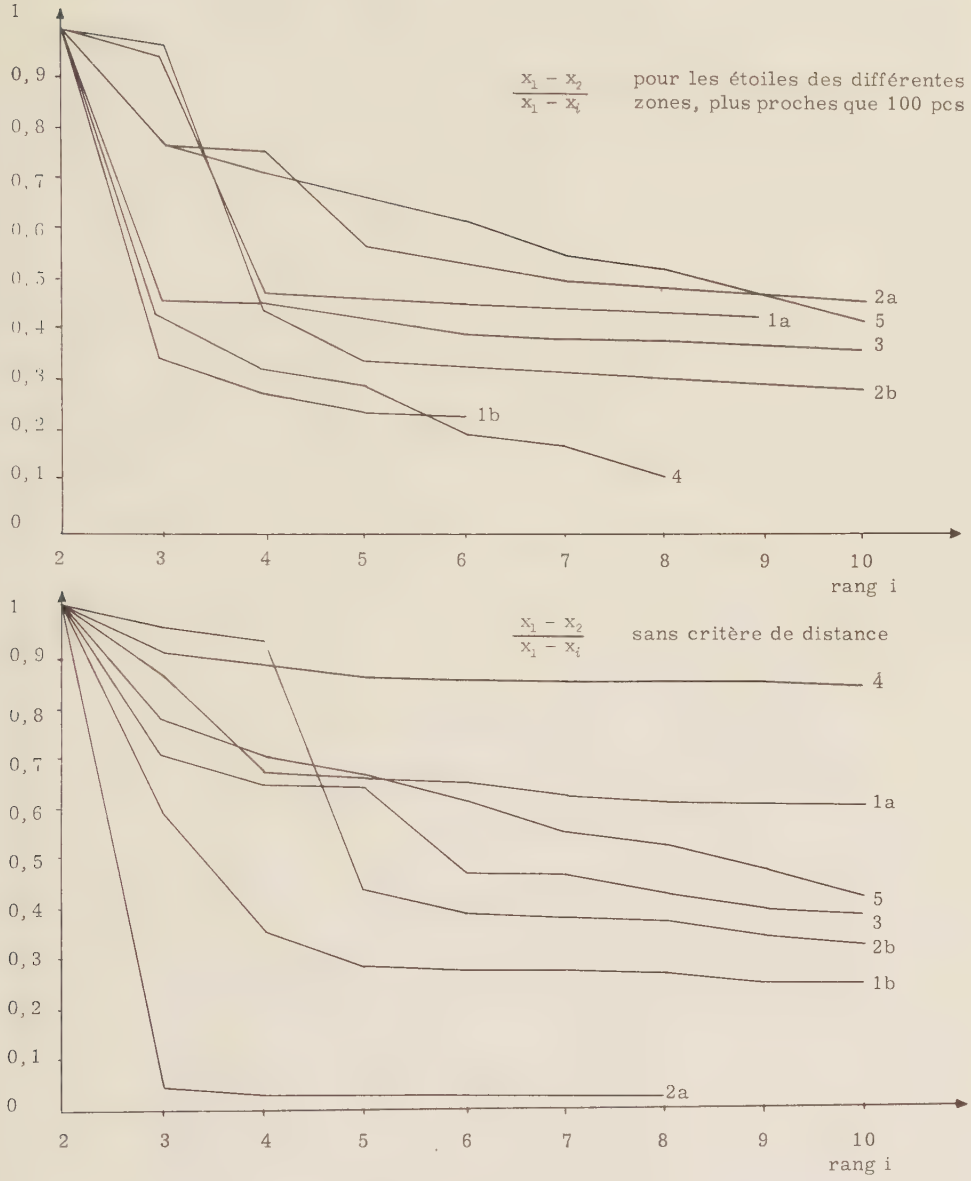
$\frac{x_1 - x_2}{x_1 - x_i}$
pour les étoiles des différentes zones, plus proches que 100 pcs
2a
5
1a
3
2b
1b
4
rang i
$\frac{x_1 - x_2}{x_1 - x_i}$
sans critère de distance
4
1a
5
3
2b
1b
2a
rang i

Figure 8 -

Par contre, il faut renoncer à l'estimation du paramètre k et nous fier à la valeur donnée par Fricke.

B - **Estimation du paramètre d'échelle b**

Ce paramètre est un paramètre d'échelle, étant donné la forme de la fonction adoptée

$$F = 1 - \left(\frac{\omega - x}{b}\right)^k$$

Autrement dit, b n'a pas de signification physique, mais présente un intérêt car, donnant l'échelle, il permet d'expliquer l'aplatissement de la courbe à proximité de la troncature dans la direction $1 = 90$, sans faire intervenir le paramètre k. Donc, dans la mesure où nous trouverons bien toujours $b = 2\omega$, notre hypothèse de garder la même valeur de k dans les différentes directions, sera justifiée.

D'autre part, b servira au calcul de la variance de ω.

Nous allons donc estimer b de la même manière que le paramètre ω.

On posera

$$\hat{b} = \sum b_i x_i$$

en imposant

$$E\,\hat{b} = b$$

et

$$\text{var } (\hat{b}) \text{ minimum}$$

La première condition conduit aux 2 équations

$$\begin{cases} \sum b_i = 0 \\ \sum \dfrac{b_i}{n^h} \dfrac{\Gamma(i+h)}{\Gamma(i)} = 1 \end{cases}$$

La deuxième condition nous oblige à procéder comme précédemment par les multiplicateurs de Lagrange.

Soit λ et μ les deux variables auxiliaires.

La fonction auxiliaire sera

$$O = \sum_i \sum_j b_i\, b_j\, C_{ij} + \lambda \sum_i b_i + \mu \left(\left(\frac{\sum b_i F_i}{n^h} \right) - 1 \right)$$

avec

$$C_{ij} = \operatorname{cov}(x_i, x_j)$$

$$C_{ii} = \operatorname{var} x_i$$

$$F_i = \frac{\Gamma(i + h)}{\Gamma(i)}$$

On a donc à résoudre le système suivant

$$\sum b_i = 0$$

$$\sum \frac{b_i}{n^h} \frac{\Gamma(i + h)}{\Gamma(i)} = 1$$

p équations $\quad \sum_j b_j C_{ij} + \lambda + \mu \frac{F_i}{n^h} = 0$

Donc $p + 2$ équations

et p inconnues $(b_i) + 2$ inconnues auxiliaires λ et μ.

Le problème est donc théoriquement résolu.
Pratiquement, nous aurons à estimer b en premier lieu.

Tableau des coefficients de l'estimateur de B.

k = 3	k = 3,5	k = 4	k = 4,5
-0,505900	-0,579558	-0,646945	-0,710749
-0,154317	-0,204070	-0,259714	-0,322968
-0,110737	-0,158412	-0,210363	-0,236370
-0,098115	-0,123224	-0,169204	-0,240829
-0,072433	-0,127216	-0,143260	-0,216575
-0,085933	-0,107284	-0,218841	-0,114334
-0,057094	-0,126625	-0,013282	-0,230926
-0,040130	-0,037658	-0,219665	-0,223464
-0,073539	-0,101593	-0,134666	+0,011505
+1,198204	+1,565643	+1,989379	+2,284713

CHAPITRE IV

INFLUENCE DES ERREURS DE MESURE SUR L'ESTIMATION

Deux méthodes s'offrent à nous : d'une part, nous pouvons faire une étude théorique, d'autre part une étude expérimentale.

A-**Correction de la distribution théorique en tenant compte des erreurs**

Le problème est classique, il s'agit de faire le produit de convolution de la fonction théorique adoptée et de la fonction d'erreur.

Pour les erreurs, nous prenons la loi de Gauss, d'écart quadratique σ. Pour la fonction théorique, nous prenons la loi de Fricke en faisant attention aux bornes d'intégration et en nous limitant au voisinage de ω où l'approximation utilisée reste valable.

Nous pouvons travailler de la manière suivante : la fonction de Gauss tend très vite vers 0 et l'on peut admettre qu'elle soit inférieure à un nombre α fixé, pour une distance à la moyenne égale à $q\sigma$ où $q \equiv q(\alpha)$.

Donc, on pourra intégrer par rapport à la variable de Gauss à partir de $-q\sigma$ en faisant la convention que $g(R)$ ne sera définie que pour $R > a + q\sigma$, alors que la fonction de Fricke approximée ne l'était que pour $X > a$.

Nous avons donc

X variable de Fricke

ε variable de Gauss

$R = X + \varepsilon$

Pour la borne supérieure d'intégration, il n'y a pas de problème, mais comme $f(x) = 0$ pour $X > \omega$, on devra avoir $R - \varepsilon \leqslant \omega$; si l'on admet que ε n'est plus intéressant au delà de $+ q\sigma$, on voit que l'on devra se limiter à

$$R \leqslant \omega - q\sigma$$

autrement dit, on va intégrer par rapport à ε de $- q\sigma$ à $+ q\sigma$ avec la convention que la variable R ne sera définie que dans l'intervalle $a + q\sigma$ à $\omega - q\sigma$.

On pourra donc écrire

$$g(R) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-q\sigma}^{+q\sigma} f(R - \varepsilon) \cdot e^{-\frac{\varepsilon^2}{2\sigma^2}} \cdot d\varepsilon$$

Soit, en prenant $f \equiv C(\omega - X)^3$

$$g(R) = \frac{C}{\sigma\sqrt{2\pi}} \int_{-q\sigma}^{+q\sigma} (\omega - R + \varepsilon)^3 \cdot e^{-\frac{\varepsilon^2}{2\sigma^2}} \cdot d\varepsilon$$

Loin de ω, on peut développer $g(R)$ et compte tenu de la symétrie de la loi de Gauss, on trouve

$$g(R) = C\{(\omega - R)^3 A + (\omega - R) B\}$$

A et B étant des constantes voisines de 1, comme $(\omega - R)^3$ est nettement plus grand, on voit que $g(R)$ est très proche de la fonction initiale.

Etudions près de ω. La borne inférieure change car on doit avoir

$$R - \varepsilon \leqslant \omega$$

or R est plus proche de ω que $q\sigma$, on ne peut plus intégrer que jusqu'à $-(\omega - R)$, on a donc

$$g(R) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-(\omega - R)}^{+q\sigma} f(R - \varepsilon) \cdot e^{-\frac{\varepsilon^2}{2\sigma^2}} \cdot d\varepsilon$$

Posons

$$\alpha = -(\omega - R)$$

$$\beta = q\sigma$$

$\varphi(x)$ pour la loi de Gauss

$\Phi(x)$ pour sa primitive

Développons le cube dans l'expression de $g(R)$, on obtient plusieurs intégrales à évaluer.

Il vient finalement avec les notations précédentes

$$g(R) = (\omega - R)^3 \ [1 - \Phi(\alpha)] + \sigma^2(\omega - R)^2 \varphi(\alpha) + \sigma^2(\omega - R) \ [1 - \Phi(\alpha)] + 2\sigma^4\varphi(\alpha)$$

Soit en revenant aux notations d'origine

$$g(R) = (\omega - R)^3 \ \Phi(x) + (\omega - R)^2 \ \varphi(x) + \sigma^2(\omega - R)\,\Phi(x) + 2\sigma^3(x)$$

avec

$$x = \frac{\omega - R}{\sigma}$$

On voit donc que l'expression obtenue est difficile à manier et, d'autre part, elle ne peut être simplifiée à priori tant que l'on ne connaît pas l'importance des différents termes.

Nous avons calculé cette expression numériquement, ce qui donne les graphiques suivants (fig. 7).

On voit que les courbes semblent décalées parallèlement à elles-mêmes avec pour $\omega = 0$ la probabilité $\frac{2 \,.\, 0,398\sigma^3}{C}$ qui donne un palier ; puis la courbe se comporte presque comme une loi de Gauss.

B - **Expérience avec des nombres au hasard**

Nous prenons la loi de Fricke et, avec une table de nombres au hasard, nous fabriquons des échantillons de variables frickiennes. Nous tirerons ensuite des nombres au hasard suivant des lois de Gauss de paramètre σ variable pour introduire des erreurs et nous chercherons à estimer le paramètre ω, quand σ varie.

Pour bien comprendre l'expérience, nous l'avons d'abord faite avec une loi uniforme.

Dans ce cas

$$g(R) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} F(R - \varepsilon)\ e^{-\frac{\varepsilon^2}{2\sigma^2}}.\ d\varepsilon$$

$$= R\,[\Phi(R) - \Phi(R - 1)] - \sigma\,[\varphi(R - 1) - \varphi(R)] + \Phi(R - 1)$$

On obtient le graphique A. (fig. 6).

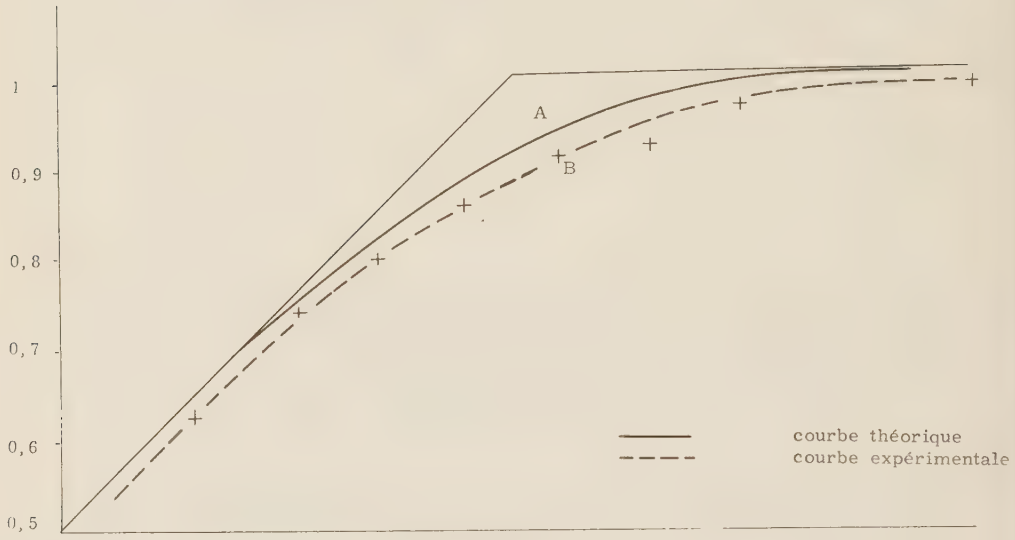

Figure 6 -

Avec des nombres au hasard, un échantillon a été fabriqué et corrigé par des erreurs (graphique B).

On trouve donc un bon accord entre la loi théorique et l'expérience.

Notons que la courbe avec les erreurs se décompose en 3 zones, l'une presque confondue avec celle de la loi uniforme initiale ; puis entre $1 - 3\sigma$ et $1 + 3\sigma$, une zone de mélange très perturbée ; puis enfin au delà de $1 + 3\sigma$, une zone qui est presque une loi de Gauss.

Figure 7 -

Nous avons ensuite fabriqué des échantillons de 1 000 valeurs, suivant la fonction de Fricke simplifiée ; soit

$$F(x) = 1 - \left(\frac{\omega - x}{b}\right)^k$$

avec les paramètres fixés aux valeurs suivantes

$$\omega = 100$$

$$b = 100$$

$$k = 4$$

Nous avons sélectionné les plus grandes valeurs seulement.

Nous nous proposons de faire des estimations avec ces 3 échantillons, puis de les corriger avec des erreurs.

1er échantillon	2e échantillon	3e échantillon
estimation $\hat{\omega} = 104$	estimation $\hat{\omega} = 87,69$	estimation $\hat{\omega} = 97,9$
x = 73,2765	x = 73,8866	x = 75,7118
x = 73,0183	x = 70,4644	x = 69,2250
x = 70,0000	x = 72,8169	x = 85,7701
x = 82,5369	x = 74,6189	x = 70,3582
x = 71,2944	x = 72,7743	x = 69,9078
x = 71,3900	x = 70,1964	x = 72,8510
x = 69,7893	x = 68,9291	x = 79,1220
x = 81,3878	x = 79,9999	x = 78,7475
x = 69,0212		
x = 68,9708		
x = 79,6351		
x = 80,2139		
x = 87,4259		

Nous voyons avec ces trois échantillons que la fluctuation de l'estimation est importante ; $\hat{\omega}$ variant de 87 à 104.

Remarquons aussi que les estimations avec la 1ère et la 5e valeur avaient donné respectivement : 100, 92, 107.

Nous avons fait d'autres expériences avec des nombres au hasard, en faisant varier b ; nous avons constaté une certaine in-

fluence du paramètre b ; comme le laisse prévoir la formule de la variance, la dispersion des estimations augmente avec b. Nous avons constaté d'autre part que, en majorité, les estimations pour un ω fixé à 100, étaient comprises entre 90 et 110 environ.

1er échantillon avec des erreurs.

Nous introduisons des erreurs ε tirées de la table de la Rand

1/ $\sigma = 1$

Nous avons fait trois tirages d'erreurs.
On trouve pour $\hat{\omega}$ successivement

- $\hat{\omega} = 106$
- $\hat{\omega} = 109$
- $\hat{\omega} = 107$

On trouve comme moyenne 107, et l'on trouve 3 d'écart avec l'estimation faite avec l'échantillon brut, soit 3 fois l'écart.

2/ $\sigma = 3$

Nous avons fait aussi trois tirages d'erreurs.
On trouve pour $\hat{\omega}$ successivement

- $\hat{\omega} = 115$
- $\hat{\omega} = 114$
- $\hat{\omega} = 111$

La moyenne est donc 113, on trouve donc encore bien une différence d'environ 3 fois l'écart.

Autrement dit, dans le cas d'erreurs, l'estimation peut donner une valeur supérieure d'environ trois fois l'écart, ce qui confirme bien la partie théorique à savoir que, approximativement, on a une fonction tronquée à une valeur $\omega + q\sigma$ au lieu de ω.

CHAPITRE V

APPLICATION A DES VITESSES RADIALES

Nous allons utiliser les méthodes mises au point dans les chapitres précédents pour estimer la vitesse d'évasion dans le plan galactique.

Rappelons que pour le paramètre ω, nous prenons comme estimateur

$$\hat{\omega} = \sum a_i x_i$$

où les a_i sont des coefficients (cf. p. 21) obtenus en imposant à l'estimation d'être correcte et à variance minimum, où les x_i sont les plus grandes valeurs de l'échantillon. Les indices i variant de 1 pour la plus grande valeur à 8 ou 10 suivant les cas.

Nous appliquons cette méthode à deux matériels de vitesses radiales

1/ des vitesses tirées du catalogue général de Wilson.

2/ des vitesses tirées du catalogue de Gliese.

Nous avons vu l'influence considérable des erreurs de mesure, on se rend donc compte qu'il faut un matériel homogène et de bonne qualité.

D'autre part, nous avons vu que la vitesse d'évasion était liée en un point au potentiel régnant. Nous nous proposons d'estimer la vitesse d'évasion au voisinage du soleil. Il nous faut donc utiliser des étoiles comprises dans un certain volume de l'espace ; donc il faut un critère de distance.

C'est pourquoi nous avons cherché à utiliser le catalogue de Gliese qui comporte des étoiles dont la distance est au plus égale à 20 parsecs. Ce catalogue est en fait un catalogue de vitesses spatiales, comprenant environ 900 étoiles. Or une fois un découpage en région effectué, on obtient dans chacune des échantillons trop petits.

Cependant si l'on admet que la répartition des vitesses est indépendante de la longitude, en projetant un nuage de vitesses spatiales sur une direction donnée, on obtient une distribution de vitesses radiales.

Nous avons utilisé ce procédé pour le catalogue de Gliese, ce qui nous fournit des échantillons d'environ 900 étoiles.

Catalogue de Gliese

Nous avons donc projeté le nuage des vitesses spatiales sur quatre directions

1 = 90 (direction de rotation)

1 = 120

1 = 150

1 = 180 (direction opposée à celle du centre galactique)

Les 4 échantillons ainsi obtenus ne sont pas rigoureusement indépendants, cependant ce ne sont pas les mêmes vitesses qui ont donné les plus grandes valeurs de chaque direction. Les vitesses sont ici données par rapport au soleil.

Résultats

Direction 1 = 90.

Dix plus grandes valeurs

58, 55, 55, 53, 49, 44, 41, 40, 39, 37.

- pour $k = \frac{9}{2}$ $\hat{\omega}_{90} = 93,5$ km/s $\sigma\omega \simeq 18$ $\hat{b}_{90} = 154$

- pour $k = \frac{7}{2}$ $\hat{\omega}_{90} = 83$ km/s $\sigma\omega \simeq 19,5$ $\hat{b}_{90} = 160$

Direction 1 = 120

Dix plus grandes valeurs

81, 78, 74, 69, 64, 58, 57, 57, 52.

- pour $k = \frac{9}{2}$ $\hat{\omega}_{120} = 127$ km/s $\sigma\omega \simeq 24$ $\hat{b}_{120} = 204$

- pour $k = \frac{7}{2}$ $\hat{\omega}_{120} = 114$ km/s $\sigma\omega \simeq 25,5$ $\hat{b}_{120} = 216$

Direction 1 = 150

Dix plus grandes valeurs

121, 115, 111, 103, 85, 83, 80, 77, 75, 73.

- pour $k = \frac{9}{2}$ $\hat{\omega}_{150} = 189$ km/s $\sigma\omega \simeq 37,5$ $\hat{b}_{150} = 326$

- pour $k = \frac{7}{2}$ $\hat{\omega}_{150} = 168$ km/s $\sigma\omega \simeq 40,5$ $\hat{b}_{150} = 272$

Direction 1 = 180

Dix plus grandes valeurs

272, 151, 130, 130, 129, 121, 120, 120.

- pour $k = \frac{9}{2}$ $\hat{\omega}_{180} = 361$ km/s $\sigma\omega \simeq 74,5$ $\hat{b}_{180} = 611$

- pour $k = \frac{7}{2}$ $\hat{\omega}_{180} = 315$ km/s $\sigma\omega \simeq 81,5$ $\hat{b}_{180} = 698$

Nous constatons que le paramètre $\hat{b}$ est sensiblement égal à deux fois $\hat{\omega}$.

Catalogue de Wilson (1)

Nous avons sélectionné dans ce catalogue des étoiles du plan galactique, en fait comprises entre $b = -30$ et $b = +30$. Pour le

(1) Des réserves ayant été faites quant au critère de sélection utilisé pour les étoiles. Cette partie du travail doit être prise comme non définitive. L'application a un matériel équivalent sera publié par la suite.

découpage en longitude, nous avons pris cinq zones. Quand cela était possible, c'est-à-dire quand il y avait assez d'étoiles, nous avons distingué les étoiles des parties symétriques ; sinon, nous regroupions les étoiles. Enfin, nous comptons les points de vitesse qui tombent dans une zone, ce qui peut correspondre soit à une vitesse positive, l'étoile étant dans la direction envisagée, soit à une vitesse négative, l'étoile étant dans la région opposée.

Les longitudes galactiques utilisées sont basées sur la direction du centre galactique ($l = 0$).

Les vitesses de la dernière colonne de nos listes sont des vitesses rapportées au centroïde de M. Delhaye (10), de coordonnées par rapport au point de vitesse du soleil $v = -12$, $u = +9$.

Nous ferons des estimations avec huit ou dix valeurs. Puis nous ferons une sélection suivant un critère de distance : nous éliminerons les étoiles plus lointaines que 100 parsecs environ ; la distance étant évaluée grossièrement à partir de la magnitude apparente et de la classe spectrale.

Vu le nombre de vitesses du catalogue de Wilson, nous avons uniquement sélectionné des queues de distribution, en évaluant le nombre approximatif d'étoiles tombant dans chaque zone.

Tableau des différentes zones.

	longitude des étoiles	signe des vitesses	points de vitesse
zone 1, a	170 et 190 ou 350 et 10	+ -	170 et 190
zone 1, b	170 et 190 ou 350 et 10	- +	350 et 10
zone 2, a	150 et 170 ou 330 et 350	+ -	150 et 170
zone 2, b	10 et 30 ou 190 et 210	+ -	10 et 30
zone 3	30 et 50 ou 210 et 230 ou 130 et 150 ou 310 et 330	+ - + -	30 et 50 ou 130 et 150
zone 4	110 et 130 ou 290 et 310 ou 50 et 70 ou 230 et 250	+ - + -	110 et 130 ou 50 et 70
zone 5	70 et 110 ou 250 et 290	+ -	70 et 110

Zone 1 ; a

N° Wilson	V_r	Mg	Spc	V_r/C	
2 420	339	8,9	dA 8	330	&(1)
4 100	195	10,0 var	cK 2 ev	186	
9 237	-174	8,0 v	gM 4 e	165	
11 040	-124	9,9	G 3	115	&
10 747	-122	.....	P d	113	
2 653	109	10,4 v	G 5	110	&
10 105	-106	9,9	P d	97	
3 635	103	6,4	M 2	94	
10 516	-103	11,3 v	R 3	94	
9 667	-102	7,6	gM 2	93	
3 645	99,9	6,4	M 0	90,9	&
9 930	98	9,3	G 3	89	&
2 627	96,3	7,8	gK 2	87,3	
3 743	96	6,8	K 0	87	&
2 280	94	12,2	dM 4	85	&
9 741	-92	6,4	K 2	83	&
4 397	86,3	6,3	F 5	77,3	&

Zone 1 ; b

N° Wilson	V_r	Mg	Spc	V_r/C	
9 989	234	8,9	F 1	242	&
11 148	198	9	G 2	207	&
10 770	175	8,5 v	Q	184	
3 662	-132	8,4 v	M 3 e	141	&
2 317	-112,5	8,3	gF 5	121,5	
10 855	109	9,3	B 0	118	
11 355	107	8,7	F 7	116	&
9 819	102	8,3	dF 3 e	111	
9 058	95	8,9 v	gM 3 e	104	
3 824	-91,5	9,3	dK 6	100,5	&
3 853	-87,1	6,4	K 0	96,1	&
2 460	-81	10,2	R 4	90	
2 221	-80	10,0 v	gM 5	89	

---

(1) étoiles à des distances < 100 pcs.

Zone 2 ; a

N° Wilson	V_r	Mg	Spc	V_r/C	
4 400	189	10,0	sdF 2	185	&
8 524	-187	.....	A 9	183	
9 723	-138	6,0	O 7 e	134	
9 205	-130	10,8 v	cK 2 ev	126	&
8 789	-130,3	9,7	P e	125,9	
2 745	112,6	7,3	dG 2	108,6	&
2 337	112	9,2	dK 4	108	&
9 712	-101	8,5	cB 0	97	
1 714	87	8,6	G 0	83	&
2 534	79,9	8,8	dF 6	75,9	&
2 978	74	9,8	dM 0	70	&
8 795	-72	10,5	dK 5	68	&
2 275	70,1	4,0	O 7 n	66,1	
8 631	-65,0	7,7	dK 5	61	&
8 713	-63	10,1	dF 4	59	&
9 680	-62,5	5,4	O 9 e	58,5	&

Zone 2 ; b

N° Wilson	V_r	Mg	Spc	V_r/C	
5 240	-240	8,2	dF 8	252	&
10 995	193,9	9,6	Pd	205	
10 404	191	9,5	dG 0	203	&
3 785	-190	9,2	dF 8	202	&
5 047	-130	3,0	G 0	142	&
10 670	119,8	10,9	Pd	132	
9 531	115	8,0	O 9	127	
3 463	-112	6,1	gM 0	124	
9 469	99,6	5,4	gM 2	112	&
10 969	95	8,9	G 5	107	&
10 298	91	6,3	A 2	103	
4 431	-86,6	8,3	dG 3	99	&
12 265	85	5,5 v	gM 5 e	97	&
11 516	78,2	9,4	dG 3	91	&
9 766	81	13,2 v	G0-G4	93	
9 764	73	7,6	G 0	85	&

Zone 3 : l compris entre 30 et 50 ou 130 et 150

Le nombre des étoiles de cette zone étant moins grand, nous ne ferons pas la décomposition en deux parties symétriques.

N° Wilson	V_r	Mg	Spc	V_r/C	
11 715	151	7,0	A 9n	166	&
11 017	112	10,6	sdF 5	127	&
11 302	96	12,0	B 9	111	
11 213	91	9,6 v	.....	106	
11 754	91,1	.....	Pc	106	
8 877	-69,1	5,7	dG 5	84	&
12 009	68	8,7	dK 5	83	&
5 231	-61,5	6,6	M 4	76	&
1 473	66,7	5,4 v	gM 4e	67,7	&
8 319	-66	10,1	B 2e	67	&
894	65,2	6,4	G 8	66,2	&
1 156	62	9,9	dM 0	63	&
8 500	-60	5,4	K 5	61	&
11 230	44	5,0 v	K 3ev	59	&
11 242	40	9,1	B 1	55	

Zone 4 : l compris entre 110 et 130 ou 50 et 70

Dans cette zone, nous ne ferons pas de subdivision.

N° Wilson	V_r	Mg	Spc	V_r/C	
11 638	194	10	WN 7	209	
12 872	71	11,5	wA	86	
10 005	58,9	7,1	gM 2	74	&
7 866	-65	5,0	F 4	71	&
261	61,1	8,2	dG 0	67	&
12 658	51	9	N	66	
11 526	50	7	K 2	65	&
12 522	50	8,3 v	gM 6	65	
12 040	50	11,6 v	cG 8v	65	
12 737	48	8,3	K 0	63	&
13 344	45	11,1	cK 2v	60	
11 389	44	5,7	gK 2	59	&
14 874	51,5	8,9	dK 2	57	&
12 051	42	11,8	R 6	57	
12 622	42,1	.....	Pd	57	
13 456	40	8,3	B 3n	55	
845	40	7,2	K 5	46	&

L'examen préalable de cet échantillon montre qu'il faut éliminer l'étoile 11 638, ceci d'une part pour des raisons statistiques

(l'écart de cette valeur avec la suivante), d'autre part, en raison du type très particulier de cette étoile et de sa distance probablement grande.

Zone 5 : l compris entre 70 et 110

Dans cette zone, nous prenons le matériel déjà utilisé dans nos premiers essais (9).

46,1 36 32,4 31 30,2 33 29,6 27,8 26,7 24,8 22

(vitesses par rapport au soleil)

Deux autres étoiles à grande vitesse sont dans cette zone, mais leur type spectral n'est pas indiqué dans les catalogues de Wilson, elles sont donc particulières et tombent dans les catégories que nous éliminons.

Avant toute estimation de ω nous voyons sur la figure 8, ce que donne la statistique $\frac{x_1 - x_2}{x_1 - x_i}$ qui devrait permettre l'estimation du paramètre k. Nous voyons comme nous l'avions annoncé, que les fluctuations sont trop grandes pour faire une estimation convenable.

Tableau des échantillons.

zone	$k = \frac{9}{2}$		$k = \frac{7}{2}$	
	échantillon brut	après élimination	échantillon brut	après élimination
1 a	468	409	432	400
1 b	*392*	*414*	351	-
2 a	346	301	299	*272*
2 b	435	403	381	*394*
2 a + 2 b	-	395	369	369
3	276	251	242	*239*
4	-	*133*	105	*114*
5	-	87	-	70

Les estimations soulignées ont été obtenues avec 8 valeurs extrêmes, les autres avec 10 valeurs extrêmes.

Vitesse d'évasion dans la direction de la rotation

Nous trouvons avec les étoiles du catalogue de Wilson 87 km/s, si l'on prend $k = 9/2$ et 70 si l'on prend $k = 7/2$; avec le catalogue de Gliese, nous trouvons 93,5 pour $k = 9/2$ et 83 km pour $k = 7/2$.

Nous trouvons donc des valeurs plus grandes avec le catalogue de Gliese.

Remarquons aussi que ces valeurs concordent bien avec celles trouvées dans une publication précédente en utilisant deux valeurs extrêmes et le même matériel du catalogue de Wilson.

Remarquons aussi que l'accord entre l'estimation du paramètre ω et du paramètre b semble meilleur pour $k = 7/2$ que pour $k = 9/2$.

Toutes ces considérations nous conduisant à une valeur de 83 km/s environ.

Cercle d'évasion dans le plan galactique

Nous avons à notre disposition des valeurs de la vitesse d'évasion dans différentes directions. Nous nous servirons de ces résultats pour ajuster le cercle d'évasion. Mais au-paravant, remarquons que la valeur trouvée dans la direction $l = 180$, fournit le rayon du cercle d'évasion et que nous pouvons trouver la vitesse circulaire au voisinage du soleil en enlevant à ce rayon, la valeur trouvée pour la vitesse d'évasion dans la direction $l = 90$.

Pour le catalogue de Gliese, cela donnerait

$$C = 268 \text{ km/s pour } k = \frac{9}{2}$$

$$C = 232 \text{ km/s pour } k = \frac{7}{2}$$

Ce sont là des valeurs assez correctes que nous pouvons préciser.

Nous allons appliquer ces résultats au calcul de la vitesse de rotation, par la méthode des moindres carrés. Remarquons, en raison de la façon dont les échantillons ont été obtenus, que cela n'est peut-être pas tout à fait justifié dans le cas du matériel de Gliese.

Rappelons rapidement la méthode

Si nous appelons C la vitesse circulaire, V_e la vitesse d'évasion par rapport au zéro des vitesses, on a

$$V_e = C \cos 1 + \omega_1$$

Si l'on utilise les estimations de ω_1, soit $\hat{\omega}_1$, on a

$$V_e = C \cos 1 + \hat{\omega}_1 + \varepsilon_1$$

ou

$$V_e - C \cos 1 - \hat{\omega}_1 = \varepsilon_1$$

Formons

$$\sum_1 \varepsilon_1^2 = \sum_1 (V_e - C \cos 1 - \hat{\omega}_1)^2$$

et nous choisissons V_e et C pour rendre minimum cette quantité.

Il vient

$$V_e \sum \cos 1 - C \sum \cos^2 1 = \sum \hat{\omega}_1 \cos 1$$

$$V_e \, n - C \sum \cos 1 = \sum \hat{\omega}_1$$

où n est le nombre de directions.

Appliquons cette méthode aux estimations obtenues avec le matériel de Gliese, après l'avoir ramené au centroïde (V = - 12, u = + 9).

Pour $k = \frac{9}{2}$ nous obtenons

$$2,37 \, V_e - 2 \, C = 314$$

$$4 \, V_e - 2,37 \, C = 777$$

qui nous donne

$$V_e = 340 \text{ km/s}$$

$$C = 246 \text{ km/s}$$

Pour $k = \frac{7}{2}$ nous obtenons

$$2,37\ V_e - 2\ C = 282$$

$$4\ V_e - 2,37\ C = 686$$

qui nous donne

$$V_e = 296 \text{ km/s}$$

$$C = 209 \text{ km/s}$$

Nous obtenons avec le matériel de Wilson avec les estimations obtenues après élimination des étoiles plus lointaines que 100 parsecs

- pour $k = \frac{9}{2}$

$$C = 323 \text{ km/s}$$

$$V_e = 431 \text{ km/s}$$

- pour $k = \frac{7}{2}$

$$C = 347 \text{ km/s}$$

$$V_e = 433 \text{ km/s}$$

Nous obtenons avec le catalogue de Wilson des valeurs beaucoup trop grandes, il y a à celà trois raisons

1/ le critère de distance utilisé est assez grossier.

2/ le matériel n'est pas homogène et les erreurs de mesures interviennent fortement.

3/ les échantillons étant peu importants, nous avons utilisé des queues de distribution trop loin de la troncature dans une zone où l'approximation n'est plus valable.

Avec le matériel de Gliese, nous trouvons au contraire des valeurs cadrant assez bien avec les constantes généralement admises pour $k = 9/2$; mais si l'on prend $k = 7/2$, comme nous avons montré qu'il était préférable de faire, nous trouvons des valeurs du rayon du cercle d'évasion et de la vitesse circulaire nettement plus petites que celles admises généralement.

En revanche, si les 250 km/s admis pour la vitesse circulaire sont certains, cela conduit à prendre $9/2$ pour k donc à trouver dans le modèle de Fricke un terme de plus bas degré d'exposant $7/2$.

CONCLUSION

Nous voici arrivés à la fin de cette étude.

Si nous admettons comme exposant du terme de plus bas degré dans le modèle de Fricke, $k = 5/2$, nous voyons que nous sommes conduits à une estimation un peu faible de la vitesse circulaire avec le matériel tiré du catalogue de Gliese, à une estimation au contraire un peu forte avec le matériel tiré du catalogue de Wilson.

Par contre, si le terme de plus bas degré a un exposant égal à $7/2$, on trouve avec le matériel de Gliese une estimation très convenable de la vitesse circulaire.

Nous voyons apparaître ainsi, dans la mesure où cette vitesse circulaire serait bien connue, un moyen d'estimer le paramètre k.

Quant aux résultats obtenus, nous attachons une plus grande signification à ceux déduits du matériel de Gliese : ceci, en raison de l'homogénéité d'une part, de la taille de ce matériel d'autre part. Nous pensons que le procédé artificiel qui a conduit à un tel échantillon est valable et nous voyons là une méthode efficace pour avoir de grands échantillons.

Nous avons fait avec des vitesses radiales de l'observatoire de Marseille, différents essais, mais les échantillons étaient trop petits pour donner de bons résultats.

Rappelons, à ce propos, que la variance de notre estimateur est de la forme

$$\frac{D}{n^{p\,2h}}$$

où D est une constante

et $h = \frac{1}{k} = \frac{2}{7}$ ou $\frac{2}{9}$

autrement dit, il faudrait un accroissement considérable des échantillons pour avoir des résultats appréciables. Mais, nous avons vu qu'il y avait deux autres facteurs intervenant

- d'une part, les erreurs de mesure ; il vaut donc mieux un matériel plus réduit, mais de très bonne qualité.

- d'autre part, la formule d'approximation à proximité de la troncature, n'étant plus valable loin de ω, il faut utiliser des vitesses proches de ω ; il faut donc une queue de distribution importante, donc des échantillons assez grands.

Et c'est là, nous semble-t-il, la difficulté de la méthode : d'une part, une estimation faite avec un grand nombre de valeurs est certainement meilleure au point de vue statistique, mais d'autre part, nous ne devons pas trop nous éloigner de la troncature.

Mais regardons une zone caractéristique à ce point de vue : la zone 1 a du catalogue de Wilson, la plus grande valeur est 330, la suivante 186. Cet interval énorme montre bien que l'on ne peut attribuer aucune valeur à cet échantillon. Or, sur la figure 8, nous voyons la zone 1 a se comporter de manière spéciale.

Peut-être pourrions-nous trouver dans cette statistique $\frac{x_1 - x_2}{x_1 - x_i}$, un critère pour juger la validité des échantillons. Il faudrait pour cela en faire une étude théorique et aussi expérimentale avec des nombres au hasard. Puis construire le graphique pour chaque échantillon et voir s'il tombe dans une zone convenable ou non. Par exemple, avec le catalogue de Wilson, nous verrions que les zones 3, 2 b, 1 b et 4 conviennent à peu près.

Nous voyons donc apparaître les conditions d'emploi de notre méthode

- échantillon important dans chaque direction.

- vitesses de bonne qualité.

- discussion de chaque échantillon.

L'intérêt de la méthode réside surtout dans la simplicité de sa mise en oeuvre.

Cependant pour la rendre plus efficace, il serait bon, outre l'amélioration du matériel, d'utiliser une formule d'approximation plus compliquée, ceci serait possible grâce à des machines électroniques.

Nous signalons que nous avons rencontré un certain nombre d'étoiles de types très particuliers qui avaient des vitesses très grandes ; or le manque d'information ne nous a pas permis de connaître leur distance. (N° Wilson 13 143 - 13 489 - 12 259). Enfin, on constate que, dans nos queues de distributions, beaucoup d'étoiles n'appartiennent pas à la séquence principale et sont souvent de type géant, de type R et Q, ou figurent avec la mention particulière.

Pour terminer, signalons que nous ne nous sommes pas préoccupés du mouvement perpendiculaire au plan galactique, mais qu'il y aurait peut-être là aussi une étude à faire sur les grandes vitesses.

BIBLIOGRAPHIE

(1) OORT - B.A.N. 159 - p. 269 - 1928.

(2) FRICKE - Astronomische Nachrichten 280 - p. 193 - 1951.

(3) GUMBEL - Statistics of extremes - Culumbia University - Press 1958.

(4) DUGUE - Statistique théorique et appliquée (analyse aléatoire) Masson 1958.

(5) CRAMER - Mathematical Methods of Statistics - Princeton U.P. 1946.

(6) BLOM - Statistical estimates and transformed Beta-variables - Stockholm 1958.

(7) GOURSAT - Cours d'analyse mathématique.

(8) F. NAHON - Comptes rendus, vol. 251 - p. 1963 - 1960.

(9) J.P. MASSONIE - Journal des observateurs vol. 44 - n° 4 - 1961.

(10) J. DELHAYE - Stars and stellar Systems vol. 5 (galactic structure) chapitre 5 (sous-presse).

(11) VON MISES - Uber die Variationsbreite einer Beofachtungsreiche. Berliner Math. Gesellschaft t. 22 p. 3 1923.

(12) GEFFROY - Contribution à la théorie des valeurs extrêmes. Publ. Institut statistique de l'Université de Paris. Vol. 7 fascicules 3 - 4. 1958.

(13) W. GLIESE - Katalog der Sterne näher als 20 Parsek für 1950,0 Astr. Rechen - Institut. Mitteilungen Serie A n° 18 Heidelberg.

(14) R.E. WILSON - General catalogue of stellar Radial velocities - 1963.

(15) Rand Corporation - A Million Random digits. Free Press.

TABLE DES MATIÈRES

	Pages
INTRODUCTION.....	1
CHAPITRE I - Problème dynamique et cinématique.....	5
CHAPITRE II - Estimation d'un paramètre de troncature...	17
CHAPITRE III - Estimation des paramètres k et b.....	31
CHAPITRE IV - Influence des erreurs de mesure sur l'estimation.....	39
CHAPITRE V - Application à des vitesses radiales.....	47
CONCLUSION.....	59

DEUXIEME THESE

Propositions données par la Faculté

"L'AXIOMATIQUE DU CALCUL DES PROBABILITES PAR KOLMOGOROV".

Vu et approuvé :
Besançon, le 9 octobre 1964
Le Doyen de la Faculté des Sciences,
R. JACQUEMAIN

Vu
et Permis d'imprimer :
Besançon, le 14 octobre 1964
Le Recteur,
Président du Conseil de l'Université
H. GAUTHIER

IMPRIMERIE LOUIS-JEAN
Ouvrages scientifiques
TYPO-OFFSET
GAP (Hautes-Alpes)
Dépôt légal n° 159
1965